Evaluation and Use of the Genetic Diversity Present in the International Cocoa Genebank (ICG,T), in Trinidad

Olivier Sounigo[2], Yvette Christopher[1], Stacy Ramdahin[1], Romina Umaharan[1] and Antoinette Sankar[1]

1 - Cocoa Research Unit, University of the West Indies, St. Augustine, Trinidad and Tobago
2 – Cirad-Cp, TA 80/02, Avenue Agropolis, 34098 Montpellier Cedex 5, France, Attached to the Cocoa Research Unit

Abstract

Genetic diversity is currently assessed in the ICG,T using Isozyme Electrophoresis and RAPD. The information obtained from this assessment is used for establishing a "working collection" within ICG,T containing a reduced number of clones with large genetic diversity and interesting agronomic traits. Proposals are made to integrate the results of this assessment in to the cocoa breeding strategy.

Introduction

The Cocoa Research Unit (CRU) is in charge of the maintenance, description and evaluation of the 2,300 cocoa clones present in the International Cocoa Genebank (ICG,T) in Trinidad. Besides the morphological description of the clones, biochemical (Isozyme Electrophoresis) and molecular (RAPD) markers are currently used to characterise these clones and to assess the level of genetic diversity present in the ICG,T.

Material and methods

Plant material

The ICG,T is composed of :

- Upper Amazon populations from Peru, Ecuador and Colombia,
- Trinitario clones from the Caribbean (Trinidad, Grenada, Martinique, Guadeloupe, Dominican Republic), Central America (Costa Rica and Mexico) and from South America (Venezuela and Colombia),
- Refractario clones from Ecuador,
- Lower Amazon Forastero from Brazil and Surinam,
- Native clones from French Guiana, and
- Some Criollo populations from Belize.

Methods

Isozyme Electrophoresis. Five systems are currently used, following the methodology of Lanaud (1986): ACP, ADH, IDH, MDH and GPI.

RAPD. DNA is extracted according to the method proposed by Edwards *et al.* (1991), as modified by Johnson *et al.* (1992). Amplifications are carried out following the method described by Christopher and Sounigo (1996). Thirteen operon primers are used to generate 30 polymorphic reproducible and scorable markers for these analyses. The amplification products are run on a 1.6% agarose gel and visualised under ultra violet light after staining with ethidium bromide.

Statistical analyses. The genetic diversity is measured using multivariate analyses. The level of genetic diversity within populations is measured by the Shannon index (Magurran 1988) when RAPD data are analysed. For isozyme data, the mean number of polymorphic markers, the mean number of alleles per locus and levels of observed and expected heterozygosity are calculated. The genetic relationships between populations are visualised using cluster analyses performed on Rogers-Wright distances (Wright 1978) in the case of RAPD data, and on Nei distances (Nei 1972) in the case of isozymes, calculated on pairs of populations.

The results and their use

Characterization using isozymes

The isozyme characterisation data for the clones are included in the International Cocoa Germplasm Database (ICGD) (Wadsworth *et al.* 1997). These can be used by all cocoa breeders for:

- Estimation of the level of heterozygosity present in the clone, even if this evaluation is not very precise (observed on only five isozyme loci). This type of information on the parents of crosses is important for the prediction of the level of heterogeneity in the progeny. This can indicate the possible future use of this progeny, as a variety (if the parents show a low level of heterozygosity) or as a reservoir for individual tree selection (if the parents are highly heterozygous).
- Verification of identity of the clones they have in common with ICG,T in order to detect mislabelling problems.

Level of genetic diversity within populations

These data are currently used for the establishment of a "working collection" for the CFC/ICCO/IPGRI project on 'Cocoa Germplasm Utilization and Conservation' (Sounigo *et al.* 2000). A sub-sample of the most useful clones in the ICG,T is selected for inclusion in this "working collection". The "usefulness" is defined in terms of resistance to diseases and favourable pod and bean characteristics. In order to maintain a high level of diversity in the working collection, an attempt was made to represent a large number of populations, the level of representation of each population depending on the level of genetic diversity it contains.

The levels of diversity observed in some of the populations using the RAPD technique are indicated in Table 1. The information about the genetic origin of the clones (number of mother trees) was obtained from the ICGD (Wadsworth *et al.*1997).The levels of genetic diversity observed are not significantly correlated to the sample sizes of these populations indicating that the differences are not simply caused by bias arising from differences in sampling. The results show that the highest levels of diversity where found in populations that supposedly originate from the lowest numbers of mother trees (SCA, IMC and MO), which is rather surprising. This could be explained in the case of MO by the hypothesis made by Bartley (1993) that genotypes from different geographic origins could have been named MO by mistake. Even the accessions from the SPA population, supposedly originating from a single pod, present a level of genetic diversity almost equal to the one shown by the NA population, which is expected to originate from 14 mother trees. The populations from French Guiana present a low level of genetic diversity, except ELP, which presents a fair level of diversity, equal to the one presented by the NA population.

Table 1. Measurement of the level of genetic diversity within 12 cocoa populations, with Shannon index (Hi) calculated on the relative frequencies of 30 RAPD markers

Population (code)	Complete name	Geographical origin	Number of analysed clones	Number of mother trees*	Hi
AMAZ	Amazonas	Peru	8	7	0.37
BORNE7	Borne 7	French Guiana	7	7	0.15
ELP	Eulepousing	French Guiana	19	10	0.27
CAM	Camopi	French Guiana	20	19	0.14
IMC	Iquitos Mixed Calabacillo	Peru	23	2#	0.35
KER	Kerindioutou	French Guiana	7	7	0.14
MO	Morona	Peru	14	1#	0.35
NA	Nanay	Peru	23	14#	0.26
PA	Parinari	Peru	21	20#	0.32
SCA	Scavina	Peru	14	1#	0.36
SPA	Seleccion Palmira	Colombia	9	1	0.27
SPEC	Specimen	Colombia	11	4	0.17

*The value refers to the likely number of mother trees which gave rise to the clones included in this analysis except for those marked # where the mother tree giving rise to each clone is unknown and the number refers to the likely number of mother trees which gave rise to the whole population.

Relationship between populations

The dendrogram (Figure 1) and the plan defined by the first two axes of a multivariate analysis (Figure 2) clearly show a separation between the populations from French Guiana and the other populations. This result shows the usefulness of cocoa types from this region for the enrichment of the cocoa germplasm collections. The unique nature of this material makes it imperative to evaluate its combining ability with parents representing other populations. This type of study has already started at Cirad French Guiana, but as yet only with clones collected on the banks of the Camopi river. This study should be completed with the evaluation of clones collected on the banks of the other rivers, especially along the Eulepousing river since this population shows the highest level of diversity.

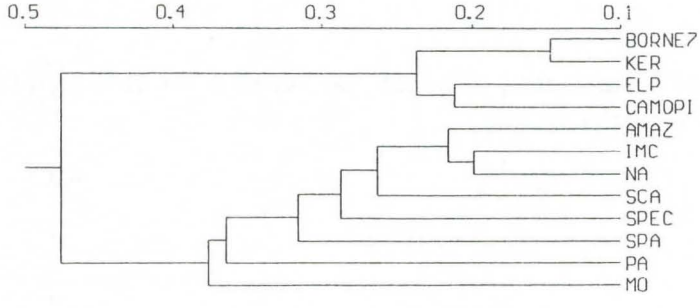

Figure 1. Dendrogram obtained from a cluster analysis performed on Rogers-Wright distances, calculated from data obtained using 30 RAPD markers, on 12 cocoa populations

Figures 1 and 3 show separation between some of the Upper Amazon Forastero populations. The results of the multivariate analysis (Figure 3) show a separation between three groups:

- LCT EEN and most of the MO clones,
- PA and some MO clones, and
- the other populations.

The dendrogram (Figure 1) shows a grouping of the IMC, NA and AMAZ populations.

These results suggest a possibility for genetic improvement based on the Upper Amazon Forastero pool by performing crosses between the following groups of populations:

- LCT EEN + MO,
- PA, and
- IMC+NA+AMAZ.

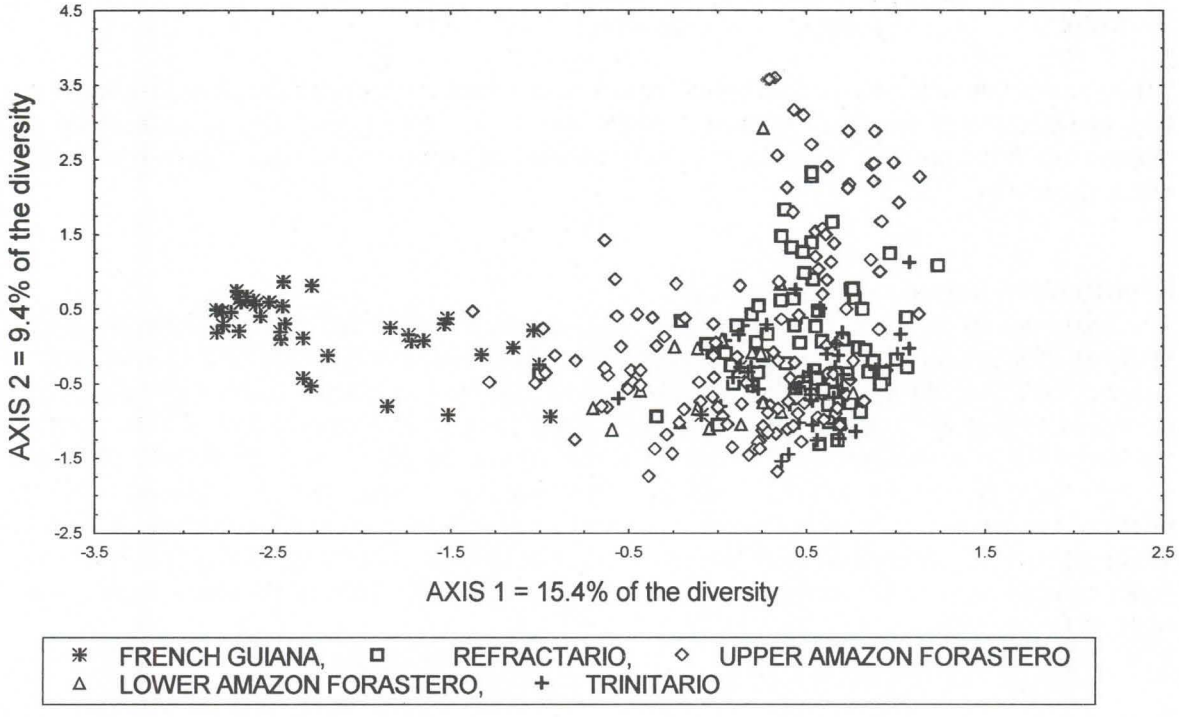

Figure 2. Principal Component Analysis performed on RAPD data obtained on 316 clones, using 29 markers

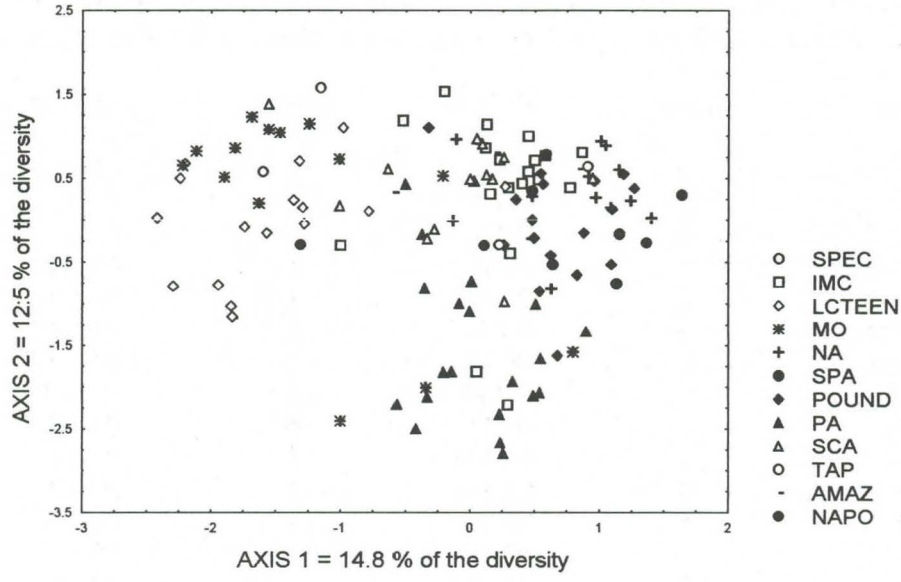

Figure 3. Plan defined by the first two axes of a Principal Component Analysis performed on data obtained on 141 clones, representing populations from Peru, Colombia and Ecuador, using 21 RAPD markers

Visualisation of the diversity using multivariate analyses

These results will assist in the process of choosing clones to include in the "working collection". The sub-sample of clones will be chosen in such a way as to maintain a high level of diversity (Figure 4 and Table 1).

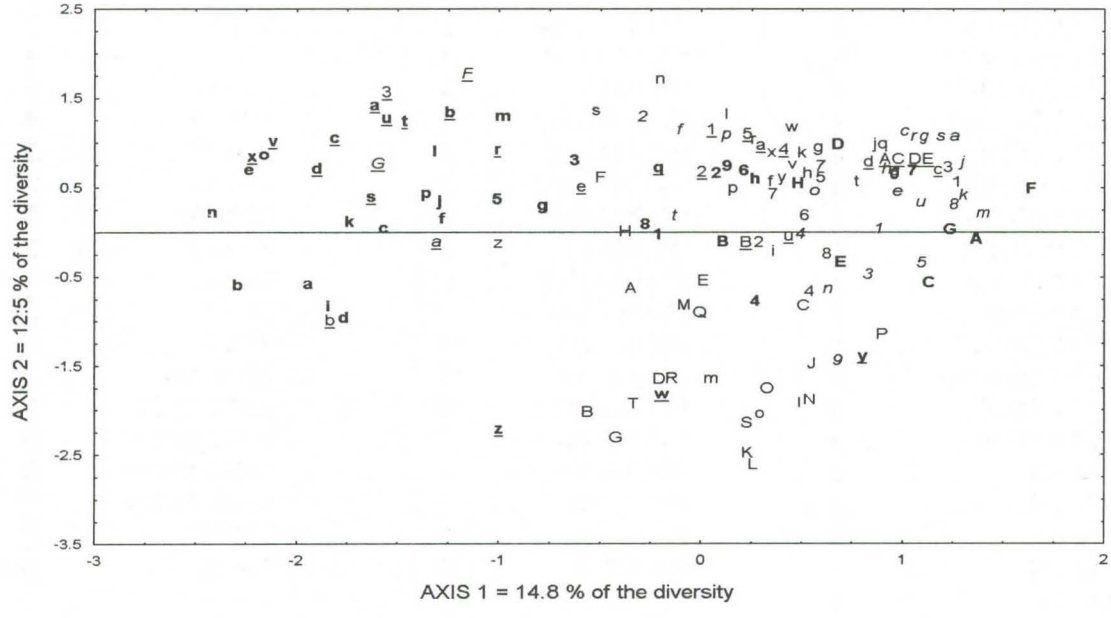

Figure 4. Principal component analysis on 141 cocoa clones from Peru, Ecuador and Colombia, using 21 RAPD markers

Table 2. List of the clones analysed and their code names in the Principal Component Analysis (see Figure 4). Clone names in bold are also analysed as reference clones in Pires *et al.* (see elsewhere in these Workshop Proceedings)

CLONE	CODE	CLONE	CODE	CLONE	CODE
AMAZ 10/1	a	MO 3	w	PA 121	D
AMAZ 11	b	MO 76	x	PA 126	E
AMAZ 12	c	MO 81	y	PA 13	F
AMAZ 12/4	d	MO 9	z	PA 149	G
AMAZ 15/15	e	MO 90	a	PA 150	H
AMAZ 5/2	u	MO 94	b	PA 165	I
IMC 105	g	MO 96	c	PA 175	J
IMC 107	h	MO 99	d	PA 184	K
IMC 11	I	NA 111	*e*	PA 188	L
IMC 14	j	NA 114	*f*	PA 194	M
IMC 16	k	NA 118	*g*	PA 200	N
IMC 3	l	NA 155	*h*	PA 211	O
IMC 31	m	NA 194	*i*	PA 30	P
IMC 38	n	NA 226	*j*	PA 300	Q
IMC 41	o	NA 26	*k*	PA 46	R
IMC 47	p	NA 370	*a*	PA 73	S
IMC 53	q	NA 43	*c*	PA 82	T
IMC 54	r	NA 46	*s*	SCA 10	**1**
IMC 57	s	NA 68	*t*	SCA 11	**2**
IMC 6	t	NA 70	*u*	SCA 12	**3**
IMC 60	f	NA 753	*m*	SCA 16	**4**
IMC 65	v	NA 79	*n*	SCA 19	**5**
IMC 67	w	NA 794	*o*	SCA 20	**6**
IMC 78	x	NA 804	*p*	SCA 23	**7**
IMC 96	y	NA 98	*r*	SCA 24	**8**
IMC 98	z	NAPO 34	*a*	SCA 27	**9**
LCT EEN 15/3	**a**	POUND 10/B	1	SCA 3	1
LCT EEN 163/D	**b**	POUND 12/A	2	SCA 5	2
LCT EEN 195	**c**	POUND 15/A	3	SCA 6	3
LCT EEN 20/S10	**d**	POUND 16/B	4	SCA 8	4
LCT EEN 201	**e**	POUND 18	5	SCA 9	5
LCT EEN 203/S3	**f**	POUND 18/A	6	SPA 10	**A**
LCT EEN 251	**g**	POUND 21/B	7	SPA 12	**B**
LCT EEN 261/S4	**h**	POUND 25/A	8	SPA 16	**C**
LCT EEN 283	**i**	POUND 26/C	9	SPA 18	**D**
LCT EEN 327	**j**	POUND 2/B	*1*	SPA 20	**E**
LCT EEN 332	**k**	POUND 31/A	*2*	SPA 4	**F**
LCT EEN 46	**l**	POUND 32/A	*3*	SPA 7	**G**
LCT EEN 6/S1	**m**	POUND 4/A	*4*	SPA 9	**H**
LCT EEN 61/S5	**n**	POUND 5/B	*5*	SPEC 4/6	A
LCT EEN 66	**o**	POUND 5/C	*6*	SPEC 41/11	B
LCT EEN 84	**p**	POUND 7/A	*7*	SPEC 41/6/17	C
MO 109	q	POUND 8/C	*8*	SPEC 41/6/18	D
MO 121	r	POUND 9/B	*9*	SPEC 41/6/25	E
MO 125	s	PA 1	A	TAP 12	*F*
MO 129	t	PA 107	B	TAP 2	*G*
MO 14	u	PA 120	C		
MO 20	v				

Conclusion

Isozyme and RAPD marker techniques currently provide us with useful information on the genetic diversity existing in the ICG,T. However, these two types of markers suffer from some drawbacks:

- dominance of the RAPD markers, which prevents information from being obtained on the level of heterozygosity of the clones, and
- low number of isozyme markers.

The introduction of another technique, especially one using co-dominant markers, would therefore be desirable. PCR-based SSR markers, developed for cocoa by Lanaud *et al.* (1999) could be a technique of choice, once the necessary equipment for sequencing gels is available at CRU.

References

Bartley B.G.D. 1993. Notes on the meaning and origins of clone names. Personal communication to Michelle End, The University of Reading, UK.

Christopher Y., O. Sounigo. 1996. The use of RAPD for characterization and genetic diversity assessment of cacao. Pages 38-51 *in* Annual report of the Cocoa Research Unit for 1995. The Cocoa Research Unit, St. Augustine, Trinidad.

Edwards K., C. Johnstone and C. Thompson. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucl. Acid. Res., 19: 1349.

Johnson E., J.R. Russell, F. Hosein, W. Powell and P. Waugh. 1992. A laboratory manual for RAPD analyses in cocoa. *In* Internal report of the Cocoa Research Unit (unpublished).

Lanaud C. 1986. Utilisation des marqueurs enzymatiques pour l'étude génétique du cacaoyer: *Theobroma cacao* L. I. Contrôle génétique et "linkage" de neuf marqueurs enzymatiques. Café Cacao Thé 30: 259-270.

Lanaud C., A.M. Risterucci, I. Pieretti, M. Falque, A. Bouet and P.J.L. Lagoda. 1999. Isolation and characterization of microsatellites in *Theobroma cacao* L. Mol. Ecol. 12 (8): 2141.

Magurran A.E. 1988. Ecological diversity and its measurement. Croom Helm. London. 179 pp.

Nei M. 1978. Genetic distance between populations. Amer. Naturalist. 106: 283-292.

Sounigo O., V. Mooleedhar, D. Iwaro, F. Bekele, T.N. Sreenivasan, J.M. Thévenin, N. Khan and D. Butler. 2000. Strategy to establish a CFC Project Collection. Pages 29-37 *in* Proceedings of the CFC/ICCO/IPGRI Project Workshop. 1-6 February 1998. Montpellier. France. IPGRI,Rome, Italy.

Wadsworth R.M., C.S. Ford, M.J. End and P. Hadley (eds). 1997. International cocoa germplasm database. The London International Financial Futures and Options Exchange (LIFFE), London, UK.

Wright S. 1978. Evolution and the genetics of populations. vol 4. Variability within and among natural populations. University of Chicago Press,Chicago, U.S.A.